

Entrega 1: Proyecto Introducción a la Inteligencia Artificial

Presentado por

Lizeth Giraldo Velez

CC: 1001237603

Angie Liseth Coronel

CC: 1085333292

Profesor

Raúl Ramos Pollan



Universidad de Antioquia
Facultad de Ingeniería

1. Planteamiento del problema predictivo

Según la OMS los accidentes de tránsito son una de las primeras causas de muerte entre la población de 15 y 29 años de edad a nivel mundial. Por lo que, en el momento que se registra un accidente de tránsito es de vital importancia conocer cuales son los riesgos potenciales que podrían conducir a un accidente, y a su vez establecer la severidad del mismo. Para este proyecto se plantea implementar un modelo de machine learning con el fin de predecir la severidad de un accidente de tránsito dependiendo del tipo y referencia de vehículo.

2. Dataset

El dataset a utilizar fue encontrado en Kaggle y contiene información detallada sobre accidentes viales en Reino Unido, con datos desde 2005 hasta 2014. Este dataset contiene 3 archivos .csv y uno .xls :

1) Un archivo que contiene la información de los accidentes con 32 columnas, llamado “*Accidents0514.csv*”, donde se dan características como condiciones climáticas, severidad del accidente, lugar, fecha, hora, día de la semana, condiciones de la carretera, latitud, longitud, límite de velocidad, etc.

2) Un archivo en el que se presentan las características de los vehículos involucrados en el accidente con 22 columnas llamado “*Vehicles0514.csv*”, que indica el tipo de automóvil, modelo, tamaño del motor, antigüedad del carro, etc.

3) Un archivo donde se recopila la información de las víctimas con 15 columnas llamado “*Casualties0514.csv*”, donde se describe la edad, sexo, clase social, etc.

4) Archivo de búsqueda llamado “*Road-Accident-Safety-Data-Guide.xls*”: En todos los archivos los datos se encuentran codificados, por lo cual en este se explica la información contenida en los otros 3 archivos.

3. Métricas de desempeño

Se plantea en un primer momento una métrica de evaluación de regresión del modelo mediante el error cuadrático medio (RMSE), con el cual se busca comparar un valor predicho con un valor observado o de referencia, donde N corresponde al número total de observaciones.

$$RMSE = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

Para este caso, se hace la aclaración de que la variable de salida “*severidad del accidente*” aun cuando es de tipo discreta, puede tener un valor en el orden de 1, 2 o

3. Como métrica de negocio se espera que el modelo ayude a los usuarios que van a adquirir un vehículo nuevo, teniendo en cuenta las mejores características asociadas a aquellas referencias que presenten un porcentaje de accidentalidad menor, de manera que prime su seguridad al volante.

4. Criterio de desempeño en producción

Lo que se espera del modelo es predecir la severidad de un accidente de tránsito dependiendo del tipo de vehículo, dicha información podría servir a los usuarios para sugerir recomendaciones a la hora de adquirir un vehículo en el mercado teniendo en cuenta los temas relacionados con accidentalidad. De manera que, si un tipo de vehículo presenta un mayor número de accidentes de alta severidad (Nivel 3) se sugiere al usuario que este tiene alto riesgo de accidentalidad.

Enlace del dataset:

<https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-variables?select=Accidents0514.csv>